DOCUMENT RESUME

ED 222 562 TM 820 720

AUTHOR McKee, Barbara G.; Manning-Curtis, Colleen
TITLE Teacher Constructed Classroom Tests: The Stepchild of Measurement Research.
PUB DATE Mar 82
NOTE 31p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, March 20-22, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Environment; Higher Education; *Measurement Techniques; Questionnaires; Research Needs; *Teacher Attitudes; *Teacher Made Tests; *Test Construction; Test Format; Test Items
IDENTIFIERS National Technical Institute for the Deaf

ABSTRACT

The results of a survey questionnaire of all teaching faculty at the National Technical Institute for the Deaf (NTID) were presented. The researchers were concerned with teacher-constructed classroom tests and the problems encountered in developing, administering, and grading them. The survey topics included: instructors' current test development practices; the format most frequently used; instructors' desire for in-service training in test development; opinions concerning areas of measurement in classroom learning in need of research; and willingness to participate in test research as well as type of contributions faculty members would make. The questionnaire and descriptive statistics for all items were provided. The survey indicated teachers used primarily self-constructed items in their tests. The short answer, or completion, question was the most frequently used item format. Paper and pencil tests were the most important component of a student's course grade. The faculty members reported a high frequency of adherence to basic test development practices. The least observed practice was having another faculty member review their tests. The demographic variable which influenced instructors' opinion scores and ratings of importance of research was their experience in a measurement course. Suggestions for future training and research were included. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222562

Teacher Constructed Classroom Tests: The Stepchild of Measurement Research

Barbara G. McKee
Assistant Professor and Research Associate

Colleen Manning-Curtis
Graduate Research Assistant

National Technical Institute for the Deaf
Rochester Institute of Technology
Rochester, New York

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B. G. McKee

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the Annual Meeting of the
National Council on Measurement in Education
New York, New York
March, 1982

TM 820 720

Abstract

This presentation and accompanying paper reports results from a survey that included all teaching faculty at the National Technical Institute for the Deaf, one of the colleges of Rochester Institute of Technology. Faculty were extensively queried on the problems encountered in developing, administering, and grading classroom tests; their current test development practices; and perceptions of their needs for in-service training. Results yielded a description of intructors' test development practices and recommendations for both professional development activities and a program of research focusing on the particular measurement problems faced by the classroom teacher.

Teacher Constructed Classroom Tests:

The Stepchild of Measurement Research

As teachers and researchers, few of us can escape the problem of assessing student learning, whether it be a unit of material we have just attempted to teach our students or a question of the efficacy of a particular instructional technique we are investigating. Sooner or later those of us involved in education are forced to face the question: Did our students learn what we intended them to learn? For a variety of reasons, most of us turn to some sort of paper and pencil device to answer this question. Further, as many times as not, we find ourselves developing our own tests because previously developed instruments do not seem to precisely fit the content or student population with which we are concerned.

In 1972 Richard Anderson wrote:

> "The thesis of this paper is that educational research workers have not yet learned how to develop achievement tests that meet the primitive first requirement for a system of measurement, namely that there is a clear and consistent definition of the things being counted." (p. 145)

Hopefully, Anderson has changed his mind during the last nine years. Certainly, the current popularity of what Bormuth (1970) calls "item writing theory" (as distinguished from item response theory) has led to a plethora of systems that tell one how to write test items. All the domain specification guidelines for criterion referenced tests are but one example. Additionally, once one has test items written, Rasch (or one of his colleagues) offer a variety of methods to analyze those items. Those of us currently working in the measurement field have little excuse for not using some type of systematic method for writing our test items and for analyzing our results once those items have been administered to sufficient numbers of subjects.

It is the contention of the authors of this paper, that few--if any--of our recently popular item development and analysis systems have filtered down to one group that writes a vast number of educational assessment measures; the classroom teachers. Within education, a great deal of controversy and research has been generated regarding the use of standardized tests. However, teacher contructed measurement devices are seldom the focus of such measurement research. In fact, these tests are so frequently ignored in the measurement literature, they can be labeled the "stepchild" of measurement research.

Any single classroom test is usually not as important to a person's academic career as are the standardized tests used for entry and placement decisions. However, in sheer time and numbers, the vast majority of tests that a student takes during their years in school will be those constructed by individual teachers for a particular class. Certainly, the cumulative importance of such classroom tests equals the importance of entry and placement tests. For example, the cumulative average of such tests form an important factor (probably the most important factor) in a students' GPA. That GPA is then often used as the criterion in making judgments about the validity of a standardized test.

Yet, for a variety of reasons, we know very little about the procedures instructors use to develop their tests. Much of course has been written about the correct way to plan and write tests or to construct certain types of items (Bloom, Hively, and Bormuth for example, all present well known but very different instructions to potential test writers). Little is known, however, about the procedures the typical instructor employs to come up with their tests.

Justifiably, the concern over the quality of tests used with deaf respondents is even greater than for the population at large. Many standardized tests have been shown to be "biased" and their results (or the interpretation

of those results) have been equivocal. It is very important that faculty obtain the training necessary to insure the tests they develop represent the latest "state of the art" in the field of assessing the classroom learning of hearing-impaired individuals. It is also important that the classroom not be ignored as a laboratory to conduct our research, for surely the whole point of most applied measurement work is to improve the accuracy of those instruments that we use to make judgments about students. One of the most important of those instruments is the teacher contructed test.

Method

Data Sources

The National Technical Institute for the Deaf (NTID) is the world's largest technical college for the deaf and is one of the ten colleges that comprise the Rochester Institute of Technology. NTID offers postsecondary educational opportunities for technical and professional careers to over 1,000 college age deaf students from all over the country. NTID also employs over 150 full time teaching faulty. It is these faculty that comprise our data base.

All faculty members at NTID were surveyed and over 50 faculty members participated in hour long interviews. Both the questionnaire and the interview covered the following topics:

a. The instructors' current tests development practices;

b. The form or format the faculty member most frequently uses for their tests or assessment measures;

c. The faculty members' desire for training in test development and their feelings about the content of such training;

d. Their opinions concerning questions that should be addressed by research in the area of measurement of classroom learning;

e. Their willingness to participate in such research and the type of contribution they are willing to make.

This paper will summarize the results of the faculty survey. The information obtained from the interviews will be presented in a later paper.

Questionnaire Format

The questionnaire consisted of four sections. The first section asked some basic demographic questions such as; sex, degree, rank, etc. (see Appendix for copy of the questionnaire). The second section asked the instructors to describe a variety of practices they follow in writing their own tests and assigning grades to students. The third section asked them to indicate their agreement (or disagreement) with a variety of statements concerning testing and the role of testing in the classroom. The fourth and final section asked the instructors to indicate how important they felt it was that research be conducted in certain areas and their willingness to work with a researcher on testing research projects. This section also asked about their interest in participating in inservice training in test development and analysis procedures.

Analysis

Descriptive statistics are provided for all the items on the questionnaire. In addition, the opinion or attitude items were clustered into two categories we have labeled, test confidence and test wariness. Descriptive information is also provided for these opinion scales. In order to examine the interrelationship of various components of the questionnaire, the following correlational analyses were also completed:

a. Pearson product correlation coefficients were computed to examine the relationship of the attitude subscale scores and demographic characteristics such as degree, rank, sex, years at NTID, etc. Additionally, the attitude subscale scores were compared to the

number of test development practices followed by the instructors and the importance of paper and pencil tests in arriving at student grades.

b. Discriminate analysis was used to determine how well the instructors' priorities for research could predict certain variables such as position prior to coming to NTID, type of courses taught, whether or not the faculty member had taken a test and measurement course etc.

c. Multiple regression was employed to examine how well the instructors' research priorities could predict such variables as the two attitude subscale scores, the willingness to work with a researcher and the number of test development practices they followed.

Results

Sample Characteristics

NTID has approximately 150 persons with faculty rank whose primary responsibility is classroom instruction. (Researchers, counselors, etc. may also hold faculty rank.) Of these 150 teaching faculty, all were sent the survey and 89 (60 percent) responded. Table 1 describes certain characteristics of our sample and, where available, compares these characteristics to the total NTID teaching faculty.

Table 1

Selected Characteristics of Faculty Responding to Questionnaire

Characteristics	Sample (Percent)	Total NTID Faculty (Percent)	Chi-Square
Department			
General Studies	7.8	11.3	
Data Processing/Business	20.2	12.7	
Engineering Technologies	18.0	11.3	
Applied Science/Applied Health	6.7	6.7	
Physics and Math	6.7	12.0	
Applied Art	11.2	9.3	
Communication	26.9	20.0	
Other	2.2	17.0	27.45 *
Rank			
Instructor	34.1	26.0	
Assistant Professor	34.1	38.0	
Associate Professor	17.0	24.0	
Full Professor	1.1	4.0	
Professional Staff	13.6	8.0	10.98 NS
Sex			
Female	46.1		
Male	53.9		
Hearing Status			
Deaf	9.1		
Hard of Hearing	4.5		
Hearing	86.4		
Degree			
Ph.D. or Ed.D.	12.6	20.5	
M.A. or M.S.	71.3	61.7	
B.A. or B.S.	13.8	11.7	
A.A.S.	2.3	3.9	5.55 NS

*p .01

In addition to those characteristics described in Table 1, our data indicate that 32 percent of the respondents came to NTID from business or industry, 10 percent from a school or college program for deaf students, 30 percent from hearing schools or colleges, and 21 percent from graduate school.

The majority of the faculty responding to the survey taught technical courses, either technical lab courses (31%) or technical classroom courses (19%). The next most common type of course taught was communications (27%), followed by General Studies (10%). This closely follows the general distribution of faculty within the NTID teaching departments. Fifty-six percent of the respondents indicated that they had taken a course in test and measurement at some time during their academic career.

Source of Test Items and Grading Practices

The degree to which teachers construct tests using self-made items is shown in Table 2. The NTID instructors indicated that they write approximately 82 percent of their test questions themselves. Of those they do not write, the majority came from other teachers' tests.

The frequency the teachers use a given item format is shown in Table 3. Short answer items were by far the most frequently used type of item and essay questions the least used. The language skills of deaf students frequently makes the scoring of essay questions extremely difficult, so this finding is not surprising.

Table 4 illustrates the way instructors indicate they arrive at a grade for their students. In spite of the problems with paper and pencil tests, it is by far the most important component of a student's grade. The numbers in Table 4 do not quite total 100 percent, because not all of the percentages listed by the instructors totaled 100.

Table 2

Sources of Test Questions Used by Teachers in their Classroom Tests

Source	Mean (percent)	S.D.
Self Constructed	82.4	27.8
Non Self-Constructed		
Tests accompanying textbook	14.6	3.6
Other teachers' tests	28.1	4.5
Items written by whole dept.	21.3	4.1
Other	36.0	--

Table 3

Rank Order of Items Formats Used by Teachers

Format	Mean Ranking*	S.D.
Short answer or completion	2.0	1.2
Multiple choice	2.4	1.2
Matching	3.0	1.2
True false	3.3	1.1
Essay	4.4	1.2

*1 = most frequently used format and 5 = least used format

Table 4

Components of Course Grades for Students

Component	Mean (percent)	S.D.
Paper and pencil tests	41.4	27.1
Papers	7.1	15.4
Class projects	14.2	20.3
Class participation	7.3	10.6
Laboratory tests	3.8	9.5
Laboratory work	10.9	16.3
Other (mostly homework)	11.6	23.0

Test Construction Principles

Of major interest for this investigation was the extent to which instructors followed any test construction principles. Table 5 summarizes the instructors' responses to the questions concerning frequency of following eleven such principles. The most frequently followed procedure was to keep a file of old items. However, the mode for several of the items was four.

Reynolds & Menard (1981) asked many of the same questions (or very similar questions) of a group of high school teachers. Table 6 compares the answers of the NTID faculty to the responses of the high school teachers. Because Reynolds used a five point scale and we used a four point scale, the responses of the NTID faculty were recoded to allow for easier comparisons (similar adjectives were assigned the same number).

Opinions Concerning Testing

Teachers were next asked to respond to a series of eighteen Likert items concerning the use and functions of paper and pencil tests. Table 7 lists those eighteen items and summarizes instructors' opinions. An exploratory factor analysis led to the development of two opinion scales, one of which we labeled test confidence, the other we labeled test wariness. The items that comprise the two subscales and the descriptive statistics concerning the subscales are presented in Table 8. Individual instructor scores were calculated for each of the subscales by summing across the included items. The subscales scores were used in subsequent analyses.

Table 5

Instructor Utilization of Test Construction Procedures

Procedure	Mean*	Mode	S.D.
Use a table of specifications	2.8	3.0	1.0
Prepare test one week in advance	2.9	3.0	0.9
Prepare answer key before administering	3.0	4.0	1.0
Follow specific rules for item writing	3.0	4.0	0.9
Develop tests so students finish in time alloted	3.6	4.0	0.7
Ask another teacher to review tests	2.0	2.0	0.8
Keep a file of previously used items	3.5	4.0	0.8
Calculate class mean	2.7	4.0	1.1
Calculate difficulty level	2.1	2.0	0.9
Provide extended time if necessary for some students to finish	2.5	4.0	1.1
Determine a passing score before administering tests	3.3	4.0	0.9

*1 = Never, 2 = Occasionally, 3 = Frequently, 4 = Always

Table 6

A Comparison of NTID Faculty and High School Teachers

Frequency of Use of Test Construction Procedures

Procedure	NTID Faculty		High School Teachers	
	$\bar{X}$	SD	$\bar{X}$	SD
Prepare table of specifications	3.68*	1.33	2.84	1.27
Prepare test one week ahead	3.87	1.06	3.25	1.17
Prepare key ahead of time	3.97	1.26	4.04	1.08
Follow item writing rules	3.94	1.18	3.80	1.03
Make test appropriate length for time allowed	4.56	.85	4.12	.81
Ask colleagues to review test	2.80	1.26	1.95	.94
Maintain item file	4.48	.94	4.15	1.26
Calculate class mean	3.55	1.5	2.90	1.27
Calculate item difficulties	2.92	1.33	2.28	1.09

*Highest possible score is 5, least possible is 1.

Table 7

Faculty Responses to Opinion Items on Testing

Item	Mean*	S.D.
Good achievement fostered by use of good tests of achievement**	3.9	1.0
Written tests not the only basis teachers should use in evaluating students	4.4	0.6
Primary function of test is to measure achievement	3.3	1.1
Good test requires teacher to define course objectives in specific, operational terms	4.3	0.8
Tests can motivate and provide learning exercises	4.4	0.7
Tests should be prepared by course instructor	4.1	0.7
A major goal of education is to develop in students command of substantive knowledge	4.0	0.8
Little known about nature and measurement of higher mental processes	3.1	0.8
Tests of moderate difficulty prove most information	3.3	0.7
Not necessary to determine a passing score	2.0	0.9
Essay tests demand student write well; objective test demand student read well	3.5	1.0
Subjective judgment involved in boy essay and objective items	3.5	0.9
True false items are efficient and convenient	2.6	1.1
Many important outcomes of education cannot be measured by paper and pencil tests	4.2	0.8
Paper and pencil tests not very useful with NTID students	1.9	0.8
Testing should give students estimate of how well they are doing	4.0	0.6
Tests useful in evaluating teachers	2.2	0.9
Too many tests given to students	2.6	0.9

*5 = Strongly Agree, 1 = Strongly Disagree
**See instructions for exact wording

Table 8
Descriptive Information Concerning Opinion Subscales

Items Included	Label	$\bar{X}$	SD	Range	Reliability
I					
Good Achievement fostered by good tests					
Tests not only basis for evaluation					
Primary function is to measure achievement	Testing Confidence	38.7	4.2	26-48	.66
Tests should be written by classroom teachers					
Education is to develop substantive knowledge					
Tests of moderate difficulty provide most information					
Essay tests require good writing, objective tests require good reading					
Paper and pencil tests not useful with NTID Students (negatively weighted)					
Testing should tell students how well they're doing					
Too many tests given (negatively weighted)					
II					
Good tests require course objective					
Test can motivate and teach					
Little known about higher mental processes					
Subjective judgement used in all tests	Testing Wariness	26.4	3.0	18-34	.56
Many outcomes can't be measured with tests					
Tests can evaluate teachers (negatively weighted)					
Too many tests given					

Priorities for Research

The faculty were asked to indicate how important they felt it was that research be conducted in certain areas related to testing. Table 9 summarizes the instructors' importance ratings. The teacher were also given the opportunity to write areas or topics that they felt needed research. Ten instructors responded, their research priorities are listed below exactly as they were written:

- Use of test results in SVP for prescriptive instruction by the teacher (SVP is a pre-college summer counseling and testing program)
- Development of standardized testing for oral communicative skills in NTID students
- Development of non standard evaluation processes
- Comparing tests with stated goals and objectives to determine appropriateness of tests
- Development of methods for helping students learn to answer short answer questions and essay questions
- Development of strategies to teach critical thinking
- Testing for minimum writing competency related to each major at NTID
- Development of teacher training in test construction and testing deaf students
- Use of standardized tests in determining entry skill levels
- Personal/social levels to determine readiness to pick a major

Table 9

Summary of NTID Instructors' Priorities of Research

Research Topic/Area	Mean rating*	S.D.
Use of standardized tests in placing NTID students in appropriate major	3.4	1.2
Relationships between item formats and student achievement	3.8	0.9
Grammar and syntax rules for paper and pencil tests to be used with NTID students	4.1	0.8
Development of non-paper and pencil testing procedures	3.8	0.9
Practical procedures for test length reduction and item selections	3.4	0.9
Development of procedures for scoring essay tests	3.0	1.1
Effects of time limitations in Testing	4.3	0.7

*5 = Very Important, 1 = Worthless

Table 10

NTID Instructors Interest in Inservice Training

Item	Response
Content of training	**Percent Interested**
Writing Objective Itesm	40%
Assessing reliability and validity of own tests	48%
Developing criterion referenced tests	29%
Alternatives to paper and pencil tests	42%
Other	3%
Length of training	
One hour lecture	32%
Two hour discussion	31%
Half-day workshop	26%
Full-day workshop	16%
Quarter long course	13%
Time of training	
During regular work day	46%
Lunch time	14%
Evening	6%
Quarter breaks	26%
Summer breaks	33%

Additional Analyses

The authors of this paper desired to go beyond a simple summary of the instructors' responses to the individual items on the questionnaire and to examine the interrelationship of some of the items. We were particularly interested in the relationship between the opinion items and the other portions of the questionnaire. Because of space limitations, we have summarized these analyses in narrative form only.

Attitudes vs. demographic characteristics

The two attitude subscales of "Testing Confidence" and "Testing Wariness" were correlated with several demographic variables. Significant results were as follows:

1. Faculty with more advance degrees were slightly more likely to have "higher testing wariness" scores ($p < .05$).
2. Deaf or hard of hearing faculty were more likely to have slightly higher "testing confidence" scores ($p < .05$).
3. Faculty who had taken a test or measurement course were slightly more likely to have higher "testing wariness" scores ($p < .05$).
4. Faculty that gave more importance to paper and pencil tests were more likely to have higher "test confidence" scores ($p < .001$) and conversely, faculty who gave less importance to paper and pencil tests in arriving at students' grades were more likely to have higher "test wariness" scores ($p < .01$).
5. There was no relationship shown between scores on the attitude subscales and faculty rank, length of time at NTID, or sex.

Attitudes vs. frequency of following test development procedures

We also calculated a total "frequency of practices" score by simply summing across the responses to the eleven testing practices items (see appendix).

There was a small positive correlation (p .05) between faculty members reported frequency of following these practices and their "testing confidence" scores.

Attitudes vs. type of courses taught

We also attempted to see if faculty attitudes differed by the type of courses they typically taught. The courses were divided into three groups roughly equivalent to liberal arts, communications and technical courses. Both the two attitude subscales and the 18 individual attitude items were used (separately) as the predictor variables in two discriminant analyses. Both analyses yielded two significant discriminant functions. Faculty who taught technical courses held similar opinions to faculty who taught liberal arts courses. Both the technical and liberal arts faculty had higher testing confidence scores than the communications people and lower testing wariness scores than the communications people. This was true whether one looked at the subscale scores or the individual opinion items.

Attitudes vs. prior position

Faculty were divided into three groups based on their positions prior to coming to NTID. The groups were business, deaf education, and hearing education. People who had come to NTID from business had the highest "test wariness" scores and those from deaf education, the lowest "test wariness" scores. Eight of the opinion items significantly entered two discriminant function coefficients that separated the three groups. In this analysis, faculty from hearing impaired education and hearing education looked relatively similar; those from business had a different pattern of scores. For example, the most important item (in the first function) was "the primary function of a classroom test is to measure student achievement." Faculty from a business background were far more likely to agree with this statement than faculty from education.

Attitudes vs. testing/measurement course

The respondents were divided into two groups based on whether they had ever taken a test and measurement course. Discriminant analysis was used to determine how well the instructors' attitudes could predict their group membership. Using the 18 individual items as predictors, one significant function was found (canonical r = .39) and it was possible to correctly classify over 67% of the subjects. Using only the two attitude subscale scores as predictors, 60 percent of the cases were correctly classified.

Attitudes vs. priorities

The two opinion subscale scores were individually regressed on the seven priority items to see how well the ratings given to each of the research priorities could predict their attitude scores. However, the R^2 for "testing wariness" subscale was .12 and for the "testing confidence" subscale was .16 neither of which were significant at the .05 level.

Priorities vs. practices, types of courses taught and prior position

Much like the opinion items, the individual importance ratings given to the research areas were used to predict the total frequency of following certain practices. This multiple regression did not yield any significant results; nor did the discriminant analysis used to determine if the priority ratings could predict the type of course usually taught. However, the discriminant analysis used to examine the relationship between the priority ratings and the instructors' position prior to coming to NTID did yield one significant function. The function was almost completely based on one priority rating. "Procedure for test length reduction and item selection." Faculty from a deaf education background gave this a very low priority rating while those from a hearing education background gave it a moderately high rating.

Priorities vs. measurement course

Interestingly, the ratings of importance given to various research areas did significantly discriminate between those faculty that had taken a measurement course and those who had not. Faculty who had taken such a course gave higher importance ratings to three of the listed areeas than those who had not. (The areas are numbered 3, 5, and 6 on the questionnaire, see appendix for wording of those items). The single significant discriminant function that emerged was able to correctly classify almost 70 percent of the cases.

Priorities vs. willingness to work with a researcher

The ratings given to several of the research areas were able to significantly predict whether or not a faculty member would be willing to work with a researcher on a testing research project. (R^2 = .26, p .01) The three priority areas that significantly entered the regression equation were numbers 2, 4, and 7 (see appendix for items).

Miscellaneous

Finally, there seems to be no relationship between desire for specific types of inservice training and type of course taught, position prior to coming to NTID, whether or not the instructor has had a course in measurement and his or her attitude subscale scores or priority ratings.

Summary and Conclusions

A reasonable cross-section of NTID faculty members responded to the questionnaire. The items concerned with the teachers' test development practices yielded the following useful information:

a. The vast majority of the items the teachers use in their tests are self constructed.

b. The most frequent type of item format used is the short answer or completion question.

c. Paper and pencil tests are by far the most important component of a student's course grade.

d. Overall, the faculty reported a relatively high frequency of following certain basic test development practices. Five of the eleven practices listed had a mode of "4" (always follow). The least followed practice was to have another instructor review their tests.

The items concerned with instructors' opinions concerning testing indicated that in general, the instructors expressed a rather positive and traditional view of paper and pencil tests. One of the items that received the highest agreement rating was "tests can help to motivate and direct student achievement, and can provide learning exercise." The teachers also strongly felt that tests provide an important basis but not the only basis teachers should use in evaluating student achievement. The instructors strongly disagreed with the statement that paper and pencil tests are not very useful with NTID students. Such a statement is frequently heard in deaf education because of the severe language problems of many deaf students. NTID instructors apparently do not agree with this concept.

areas. When instructors were asked to rate the importance of conducting research help students understand their strengths and weaknesses." The area rated as least important was the "development of standardized procedures for scoring essay tests." Approximately half the instructors indicated they would be willing to work with a member of the research staff on a testing research project. Nineteen percent said they would not be willing, the remainder said it would depend on the circumstances.

When instructors were queried about their desire for inservice training in the area of test development and test analysis, 79 percent indicated they would be interested in participating in such training in one or more areas. The most popular of the areas listed was "assessing the reliability and validity of your own tests." One hour lectures or two hour discussions were the most frequently chosen length or type of training. There was a perfect linear relationship between amount of time and the number of people that chose that option. (The shorter the time involved, the more people who chose that option.) Most people also preferred to have such training offered during the regular work day and, understandably, the least popular choice of times was the evening.

The demographic variable that seemed to make the most difference in the instructors' opinion scores and their ratings of the importance of certain research areas was whether or not they had ever taken a measurement course. This variable appeared to be more important than degree status, length of time at NTID and position prior to coming to NTID.

Suggestions for Future Training and Research

Training

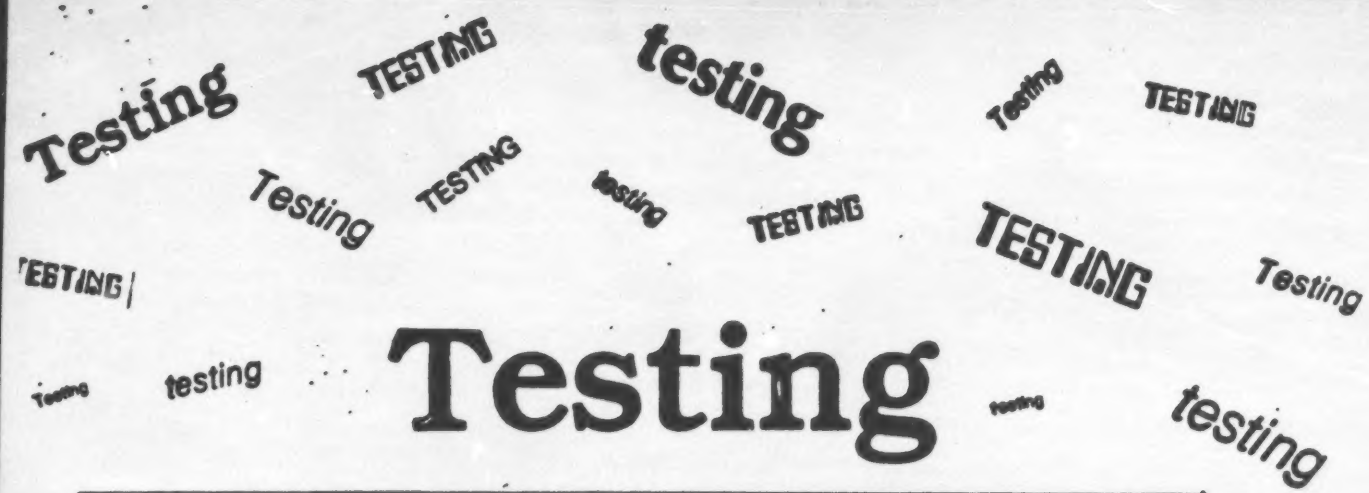
1. Focus inservice training on methods of improving tests that teachers can use quickly and easily.
2. The majority of teachers are more interested in writing better objective items than in writing (or grading) essay items.
3. Teachers are interested in learning more about alternatives to paper and pencil test items.
4. Keep sessions short, offer several one or two hour sessions rather than one all day session.
5. Offer sessions during the regular work day.

Research

1. Focus on use of tests as diagnostic tools as well as assessment tools.
2. Help teachers develop list of do's and don'ts for writing test items.
3. There is a strong interest in alternatives to paper and pencil tests.
4. More emphasis should be placed on the linguistics involved in writing tests for NTID students.
5. There seems to be a strong need to gather together in one document what we know about writing classroom tests for NTID students.

A five-year research and inservice training program on classroom assessment has been developed based, in part, on this survey. That plan is available from the first author.

B.McKee-21#
R(2/11/82):df



The Department of Educational Research and Development has need of some baseline information from NTID instructors on classroom tests. Please take a few minutes to answer this questionnaire. (It should require only about 10 minutes.) Return the completed forms as soon as possible to:

Barbara McKee, 50-1129
Educational Research & Development

Your department?_____

Your academic rank?_____
(or title, if professional staff)

Highest degree received?

- Ph.D. or Ed.D ____
- M.A. or M.S. ____
- B.A. or B.S. ____
- A.A.S. ____

Your position prior to coming to NTID?

- Business/industry ____
- School for the deaf ____
- College or college program for the deaf ____
- Hearing school ____
- Hearing college ____
- Was in college myself ____
- Other (explain) _____

_____

Your sex?

- Male ____
- Female ____

Are you?

- Deaf ____
- Hard of hearing ____
- Hearing ____

How would you describe the type of course(s) you usually teach?

- General education courses ____
- Developmental education ____
- Communication courses ____
- Non technical lab course (art, etc.) ____
- Technical lab course (engineering, etc.) ____
- Technical classroom course ____
- Individualized instruction (MLC, PLC, etc.) ____
- Other(explain)_____

_____

Years you have worked at NTID? ____

Have you ever taken a course in testing or measurement?

- Yes ____
- No ____

PRACTICES

1. What percent of the test questions that you use for your paper and pencil tests do you write yourself?

   _____

2. For those test questions you do not write yourself, where do you usually obtain those items?

   Tests accompanying textbook _____
   Other teachers' tests _____
   Items written by whole department _____
   Other (explain) _____

   _____

3. Please indicate how you typically arrive at a grade for a student by giving a percent to each of the following (they should total 100%).

   Paper and pencil tests _____
   Papers _____
   Class projects _____
   Class participation _____
   Laboratory tests _____
   Laboratory work _____
   Other (explain) _____

   _____

4. For your paper and pencil tests, please rank order the following types of items from 1 (most used) to 5 (least used) to indicate the extenst you use each.

   Essay _____
   Short answer or completion _____
   Matching _____
   Multiple choice _____
   True false _____

Please indicate the frequency you do the following by circling the appropriate letter.

	Always	Frequently	Occasionally	Never
5. When I write my own classroom tests I prepare a specific plan (table of specifications) for relating the content to be tested with different kinds of test questions.	A	F	O	N
6. I prepare my own classroom tests at least one week in advance.	A	F	O	N
7. I prepare an answer key for my tests before I administer the test.	A	F	O	N
8. When I write my own tests, I follow specific rules for item writing (e.g., avoid the use of words such as 'always' and 'never' in true false items).	A	F	O	N
9. I try to write my own tests so that all the students finish in the time available.	A	F	O	N
10. I ask another teacher to review or edit the tests I write.	A	F	O	N
11. I keep a file of items I have used on previous tests.	A	F	O	N
12. I calculate the class mean (average) for my tests.	A	F	O	N
13. I calculate the percentage of students who get each item correct (difficulty level) for my tests.	A	F	O	N
14. When a student does not finish a test on time, I allow him or her to stay after class or finish the test at a later time.	A	F	O	N
15. I determine a passing score before I administer a test	A	F	O	N

OPINIONS

Please indicate your agreement or disagreement with the following statements by circling the appropriate letter.

	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagr
1. Good achievement is fostered by the use of good tests of educational achievement.	S	A	N	D	SD
2. Written tests provide an important basis but not the only basis teachers should use in evaluating student achievement.	S	A	N	D	SD
3. The primary function of a classroom test is to measure student achievement.	S	A	N	D	SD
4. The development of a good classroom test requires the teacher to define the course objectives in specific, operational terms.	S	A	N	D	SD
5. Classroom tests can help to motivate and direct student achievement, and can provide learning exercises.	S	A	N	D	SD
6. Most classroom tests are, and ought to be, prepared by the course instructor.	S	A	N	D	SD
7. A major goal of education is to develop in the student a command of substantive knowledge.	S	A	N	D	SD
8. Little is known at present about the nature and measurement of so-called higher mental processes.	S	A	N	D	SD
9. In most tests of achievement, items of moderate difficulty, which are answered correctly by 40 to 70% of the students, contribute the greatest amount of useful information.	S	A	N	D	SD
10. Determination of a passing score on classroom tests is ordinarily unnecessary and undesirable.	S	A	N	D	SD
11. Essay tests demand that a student write well; objective tests demand that he read well.	S	A	N	D	SD
12. Subjective judgment is involved in both essay and objective testing.	S	A	N	D	SD
13. True false items are efficient and convient and can provide direct tests of a students' command of essential knowledge.	S	A	N	D	SD
14. Many of the important outcomes of education cannot be measured by paper and pencil tests.	S	A	N	D	SD
15. Paper and pencil tests are not very useful with NTID students.	S	A	N	D	SD
16. Through testing, students should have a fairly accurate estimate of how well they are doing.	S	A	N	D	SD
17. Tests are the best way of evaluating a teachers' effectiveness.	S	A	N	D	SD
18. Far too many tests are given to students.	S	A	N	D	SD

PRIORITIES AND TRAINING

Listed below are several areas or topics related to testing of NTID students. Please indicate how important you feel it is that research be conducted in a given area by circling the appropriate letter.

	Very Important	Important	So-So	Not Important	Worthless
1. The use of standardized tests in placing NTID students in appropriate majors.	V	I	S	N	W
2. The relationships between different item formats and students correct answers to test questions.	V	I	S	N	W
3. The identification of grammar and syntax rules for use in writing paper and pencil tests for NTID students.	V	I	S	N	W
4. The development and validation of non-paper and pencil testing procedures.	V	I	S	N	W
5. Practical procedures for test length reduction and item selection.	V	I	S	N	W
6. Development of a standardized procedure for scoring essay tests.	V	I	S	N	W
7. The effects of time limitation in testing NTID students.	V	I	S	N	W

8. Other (write your own) _____

_____

9. Would you be willing to work with a researcher on a testing research project; for example, modifying one of your classroom tests to examine various item formats or trying some alternative to paper and pencil tests?

___Yes ___No ___Perhaps (explain)_____

10. Below are some topics that could serve as the basis for inservice training on testing. Please check any topics you would be interested in learning more about.

___Writing objective test items
___Assessing the reliability and validity of your own tests
___Developing criterion referenced tests
___Alternatives to paper and pencil tests
___Other (explain)_____

_____

11. Below are some times for inservice workshops. If you are interested, please check the amount of time you feel you could devote to such training and when such training would be most feasible for you.

___One hour lecture
___Two hour discussion
___Half-day workshop
___Full-day workshop
___Quarter long course (20-30 hrs.)

___During regular work day
___Lunch hours
___Evenings
___Quarter breaks
___Summer breaks

PLEASE USE BACK OF PAPER FOR ANY ADDITIONAL COMMENTS

Return as soon as possible to:
Barbara McKee 50-1129
Educational Research & Development

PRIORITIES AND TRAINING

Listed below are several areas or topics related to testing of NTID students. Please indicate how important you feel it is that research be conducted in a given area by circling the appropriate letter.

	Very Important	Important	So-So	Not Important	Worthless
1. The use of standardized tests in placing NTID students in appropriate majors.	V	I	S	N	W
2. The relationships between different item formats and students correct answers to test questions.	V	I	S	N	W
3. The identification of grammar and syntax rules for use in writing paper and pencil tests for NTID students.	V	I	S	N	W
4. The development and validation of non-paper and pencil testing procedures.	V	I	S	N	W
5. Practical procedures for test length reduction and item selection.	V	I	S	N	W
6. Development of a standardized procedure for scoring essay tests.	V	I	S	N	W
7. The effects of time limitation in testing NTID students.	V	I	S	N	W

8. Other (write your own) _____

_____

9. Would you be willing to work with a researcher on a testing research project; for example, modifying one of your classroom tests to examine various item formats or trying some alternative to paper and pencil tests?

___Yes ___No ___Perhaps (explain)_____

10. Below are some topics that could serve as the basis for inservice training on testing. Please check any topics you would be interested in learning more about.

___Writing objective test items
___Assessing the reliability and validity of your own tests
___Developing criterion referenced tests
___Alternatives to paper and pencil tests
___Other (explain)_____

_____

11. Below are some times for inservice workshops. If you are interested, please check the amount of time you feel you could devote to such training and when such training would be most feasible for you.

___One hour lecture
___Two hour discussion
___Half-day workshop
___Full-day workshop
___Quarter long course (20-30 hrs.)

___During regular work day
___Lunch hours
___Evenings
___Quarter breaks
___Summer breaks

PLEASE USE BACK OF PAPER FOR ANY ADDITIONAL COMMENTS

Return as soon as possible to:
Barbara McKee 50-1129
Educational Research & Development